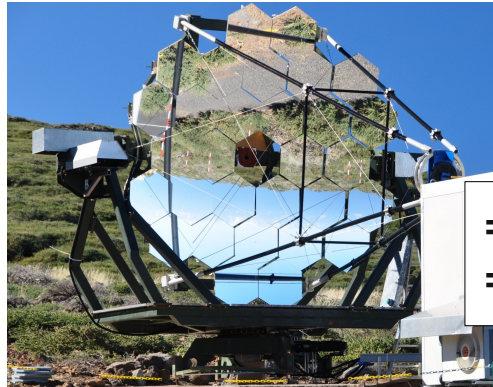


# Simulations and Data Processing

- Data flow
- Data products
- Simulation products
- Status

# Data Flow



=> Raw data  
=> Auxiliary data

Raw data <=

Simulation:  
Air Showers  
& Detector  
(gammas)

Calibration

Calibrated Data

Calibrated Data

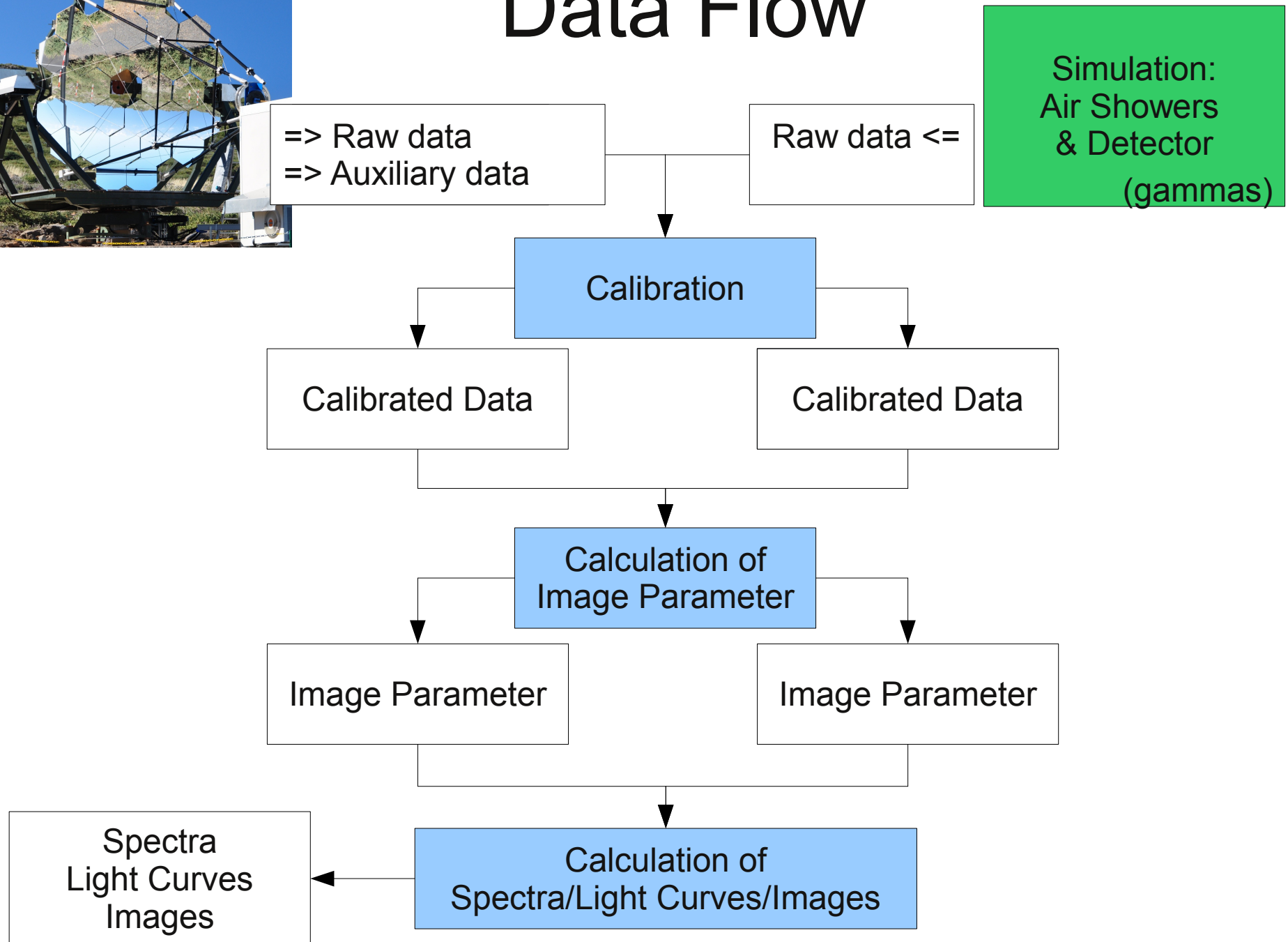
Calculation of  
Image Parameter

Image Parameter

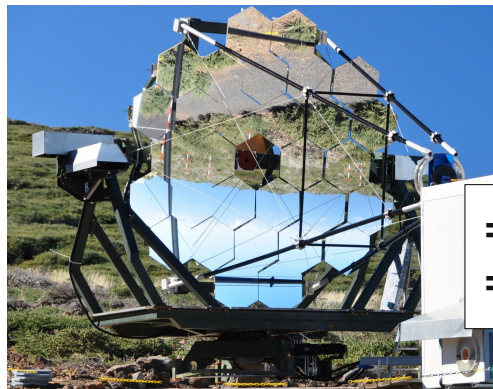
Image Parameter

Spectra  
Light Curves  
Images

Calculation of  
Spectra/Light Curves/Images



# Data Flow



=> Raw data  
=> Auxiliary data

Raw data <=

Simulation:  
Air Showers  
& Detector  
(gammas+protons)

Calibration

Calibrated Data

Calibrated Data

Calculation of  
Image Parameter

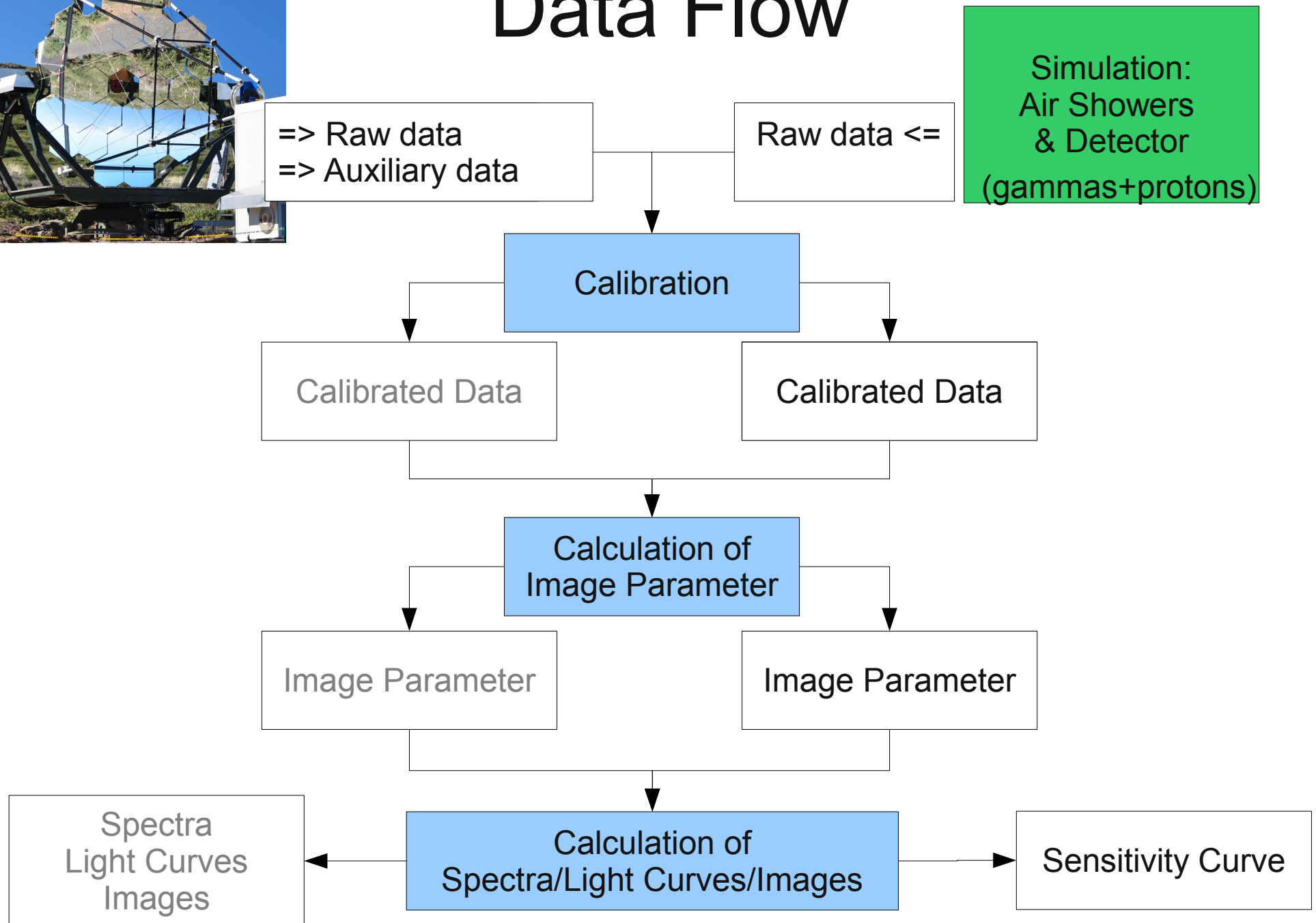
Image Parameter

Image Parameter

Spectra  
Light Curves  
Images

Calculation of  
Spectra/Light Curves/Images

Sensitivity Curve



# Status

- Automatic production, processing of simulated data

## produced statistics:

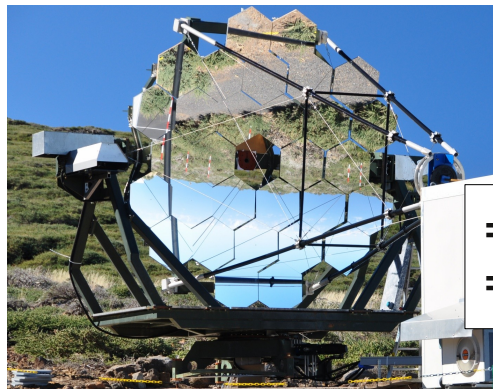
- Gammas:  $9 \cdot 10^6$  showers ( $\sim 50000$  evts at image parameter level)
- Protons:  $3.3 \cdot 10^8$  showers (  $\sim 2000$  evts at image parameter level)
- Myons:  $3 \cdot 10^8$  showers (  $\sim 9000$  evts at image parameter level)

more proton statistics currently produced in Dortmund

# Status

- Automatic production, processing of simulated data
- Preliminary sensitivity curve (not yet optimized)
- Processing up to image parameters automatic

# Processing Status



=> Raw data  
=> Auxiliary data

Raw data <=

Simulation:  
Air Showers  
& Detector  
(gammas+protons)

already implemented  
for FACT / work ongoing

Calibration

Calibrated Data

Calibrated Data

Calculation of  
Image Parameter

existing from MAGIC

Image Parameter

Image Parameter

work ongoing

Spectra  
Light Curves  
Images

Calculation of  
Spectra/Light Curves/Images

Sensitivity Curve

# Status

- Automatic production, processing of simulated data
- Preliminary sensitivity curve (not yet optimized)
- Processing up to image parameters automatic
- Still to be done:
  - Monte Carlo on demand (tailored MC for data)
  - Automatic data check and data set building
  - Automatic production of spectra, light curves and images

# Data Products

Real Data

Simulations



Simulated showers



Raw data (L0)



Calibrated data (L1)



Image parameters (L2)



Spectra, light curves, images (L3) (✓)



Auxiliary data



Processing log files



for data

for sensitivity



# Data Amount per Year

## Real Data

## Simulations



Simulated showers

✓ O(50TB) ✓ 5 TB

50 TB  
(compressed)



Raw data (L0)

✓ O(10TB) ✓ 1 TB

7 TB

different setups

10 TB



Calibrated data (L1)

✓ O(500GB) ✓ 50 GB

350 GB

different setups

500 GB



Image parameters (L2)

✓ O(25GB) ✓ 25 GB

175 GB

different setups

500 GB



Spectra, light curves, images (L3) (✓)

✓ <10 GB

1 TB



Auxiliary data



Processing log files



< 10 GB  
(for all processing)

for data

for sensitivity

(only rough estimates)

Numbers calculated assuming root files

# Size per File

## Real Data



O(1-10GB) ✓

O(1GB) ✓

O(10MB) ✓

O(MB-GB) ✓

O(1-100MB) ✓

O(KB-MB) ✓

Simulated showers

Raw data (L0)

Calibrated data (L1)

Image parameters (L2)

Spectra, light curves, images (L3) (✓) O(MB-GB) ✓

Auxiliary data

Processing log files

## Simulations

✓ O(GB) ✓

✓ O(100MB) ✓

✓ O(10MB) ✓

✓ O(1MB) ✓

✗ ✗

✓ O(KB-MB) ✓

for data for sensitivity

# Number of Files per Year

Real Data

Simulations



Simulated showers

✓  $O(20000)$  ✓  $\sim 5000$

$O(20000)$  ✓

Raw data (L0)

✓  $O(20000)$  ✓  $\sim 5000$

$O(20000)$  ✓

Calibrated data (L1)

✓  $O(20000)$  ✓  $\sim 5000$

$O(20000)$  ✓

Image parameters (L2)

✓  $O(20000)$  ✓  $\sim 5000$

$O(30000)$  ✓

Spectra, light curves, images (L3) (✓)  $O(30000)$  ✓  $\sim 1000$

$O(20000)$  ✓

Auxiliary data



$O(300000)$  in case of FITS  
(one file per subsystem)

$O(20000)$  ✓

Processing log files



for data

for sensitivity

# Possibilities to Save Storage Space

- Removing CORSIKA files after 1-2 years (given that enough CPUs are available)
- Removing unphysical events (0-20% of raw data) after 1-2 years
- Removing proton events (0-20% of raw data) after 2 years
- Clever lossless compression algorithm for raw data

# Summary

- About 100 TB per year
  - Roughly half data and half simulations
  - About 80% bulk data (raw data) and 20% output
  - About 10% end results
- About 250 000 – 500 000 files per year
  - O(GB) 50 000 – 100 000 files
  - O(10-100MB) 50 000 – 100 000 files
  - O(KB-MB) 120 000 – 400 000 files